

## CLAIMS

5 1. A method of detecting a facial region within a video comprising the steps of:

10 (a) receiving a first frame of said video comprising a plurality of pixels;

15 (b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference image based on a transform of said difference image in a spacial domain to a parameter space; and

20 (e) fitting said plurality of candidate facial regions to said difference image to select one of said candidate facial regions.

25 2. The method of claim 1 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

30 3. The method of claim 2 wherein said threshold value is a predetermined value and said removing values is setting said values of said difference image that are less than said threshold value to a selected value.

4. The method of claim 1 wherein said transform is a Hough transform.

5. The method of claim 4 wherein said Hough transform is

$$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \quad \forall x_c, y_c, r \in (x - x_c)^2 + (y - y_c)^2 = r^2.$$

5           6. The method of claim 1 where said fitting of each of said candidate facial regions is based on a combination of at least three factors including, a fit factor representative of a fit of said candidate facial regions to said difference image, a location factor 10 representative of the location of said candidate facial regions within said video, and a size factor representative of the size of said candidate facial regions.

15           7. The method of claim 1 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

20           8. The method of claim 1 wherein said step of determining said plurality of said candidate facial regions and fitting said plurality of said candidate facial regions further comprises the steps of:

25           (a) determining a set of candidate circles based on a Hough transform of said difference image;

30           (b) scoring said set of said candidate circles based on a combination of at least three factors including, a fit factor representative of the fit of said candidate circles to said difference image, a location factor representative of the location of said candidate circles within said video, and a size factor representative of the size of said candidate circles;

5 (c) selecting at least one of said candidate circles based on said scoring;

10 (d) generating at least one candidate facial region having an elliptical shape for each of said at least one of said candidate circles; and

15 (e) scoring each of said candidate facial regions based on a combination of at least three factors including, a fit factor representative of the fit of a respective said candidate facial region to said difference image, a location factor representative of the location of said respective said candidate facial region within said video, and a size factor representative of the size of said respective said candidate facial region.

20 9. The method of claim 8 wherein said generating at least one candidate facial region has a center of said elliptical shape located within a bounded region of potential locations having a greater vertical dimension than a horizontal dimension centered about the center of said respective said candidate circle.

25 10. A method of detecting a facial region within a video comprising the steps of:

30 (a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

35 (c) calculating a difference frame representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

20  
COT 5

10

(d) determining a plurality of candidate facial regions within said difference frame; and

(e) fitting said candidate facial regions to said difference image to select one of said candidate facial regions based on a combination of at least two of the following three factors including, a fit factor representative of the fit of said candidate facial regions to said difference image, a location factor representative of the location of said candidate facial regions within said video, and a size factor representative of the size of said candidate facial regions.

15

11. The method of claim 10 where said determining said candidate facial regions is based on a Hough transform of said difference image in a spacial domain to a parameter space.

20  
12. The method of claim 11 wherein said Hough transform is

$$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \quad \forall x_c, y_c, r \in (x - x_c)^2 + (y - y_c)^2 = r^2.$$

25  
13. The method of claim 10 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

30  
14. The method of claim 10 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

15. A method of determining sensitivity information for a video comprising the steps of:

- (a) receiving a first frame of said video;
- (b) receiving a subsequent frame of said video;
- (c) determining a location of a facial region within said video based on said first and subsequent frames; and
- (d) calculating a sensitivity value for each of a plurality of locations within said video based upon both said location of said facial region within said video in relation to said plurality of locations and a non-linear model of the sensitivity of a human visual system.

16. The method of claim 15 wherein the step of said calculating said sensitivity values is further based upon calculating an eccentricity versus image location in relation to a viewer of said video for said plurality of locations within said video.

17. The method of claim 16 wherein said calculating said sensitivity is further based upon a sensitivity versus eccentricity non-linear model of said human visual system.

18. The method of claim 16 wherein said eccentricity is derived according to the following,

5 where  $\theta_E$  is said eccentricity,  $y$  is a vertical pixel position within said video,  $x$  is a horizontal position within said video,  $x_c$  represents a horizontal component of a center position of an elliptical said facial region,  $y_c$  represents a vertical component of said center position of said elliptical said facial region,  $x_r$  represents a first elliptical radii of said elliptical said facial feature in a horizontal direction;  $y_r$  represents a second elliptical radii of said elliptical said facial feature in a vertical direction, and  $v$  represents a viewing distance.

10

15 19. The method of claim 15 wherein said sensitivity values are based upon the distance from the outer edge of said facial region to said plurality of locations within said video.

20

25 20. The method of claim 17 wherein said sensitivity versus eccentricity non-linear model is derived according to the following,

where  $S$  is representative of said sensitivity,  $k_{ECC}$  is a constant, and  $\theta_E$  is representative of a non-linear contrast sensitivity function.

21. A method of encoding a video comprising the steps of:

30 (a) receiving a frame of said video consisting of a plurality of pixels;

35 (b) calculating sensitivity information for a plurality of locations within said video calculated based upon the sensitivity of a human visual system of a viewer observing a particular region of said video; and

(c) encoding said frame in a manner that provides a substantially uniform apparent quality of said plurality of locations to said viewer when said viewer is observing said particular region of said video.

5 22. The method of claim 21 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

10 23. The method of claim 22 wherein said encoding is derived in accordance with the following:  
15  $Q/S_1, Q/S_2, Q/S_3, \dots, Q/S_N$   
where  $Q$  is representative of said base quantization factor, and  $S_1$  through  $S_N$  are representative of said sensitivity information for said plurality of locations.

20 24. The method of claim 23 wherein one of said  $S_k$ , where  $k$  is a value from 1 to  $N$ , is derived based upon a statistical calculation of a plurality of said sensitivity information for one of said locations of said image.

25 25. The method of claim 24 wherein  $S_k$  is an average of said plurality of said sensitivity information.

26. The method of claim 24 wherein  $S_k$  is a maximum of said plurality of said sensitivity information.

27. The method of claim 21 wherein said encoding said frame results in the total number of bits produced for said frame being substantially equal to a preselected number.

5

28. The method of claim 27 wherein said frame is encoded only once.

10 29. The method of claim 27 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

15 30. The method of claim 29 wherein said base quantization factor is derived in accordance with the following:

20 where  $A$  is representative of the number of pixels in one of said plurality of locations,  $K$  and  $C$  are constants associated with said plurality of locations,  $N$  is representative of the number of said plurality of locations, 25  $B$  is representative of said total number of bits, the  $\sigma_i^2$  values are a measure how much texture is associated with said plurality of locations, and the  $S_i^2$  values are representative of the respective said sensitivity information squared.

30

31. A method for encoding multiple blocks in a frame of image data, comprising:

(a) identifying a target bit value equal to a total number of bits available for encoding the frame;

35

- (b) calculating sensitivity information for each one of the blocks based upon the sensitivity of a human visual system observing a particular region of the image;
- (c) adapting quantization values for each of the multiple blocks to provide substantially uniform apparent quality of each of the blocks in the frame subject to a constraint that the total number of bits available for encoding the frame is equal to the target bit value; and
- (d) encoding the blocks with the quantization values.

32. The method of claim 31 wherein the quantization values are derived according to the following,

20 where,  $Q_i$  is the quantization value for each block  $i$ ,  $N$  is  
the number of blocks in the frame,  $B$  is the total number of  
bits available for encoding the frame,  $A$  is a number of  
pixels in each of the multiple blocks,  $K$  and  $C$  are constants  
associated with the image blocks,  $\sigma_i$  is an empirical  
25 standard deviation of pixel values in the block, and  $S_i$  is a  
weighting incorporating the sensitivity information for the  
block.

33. The method of claim 31 including adjusting  
30 the quantization values according to a number of image  
blocks remaining to be encoded, a number of bits still  
available for encoding the remaining image blocks, and a  
value that depends on the sensitivity and texture of the  
remaining image blocks.

34. The method of claim 32 including using a K parameter and a C parameter on a block-by-block basis to adjust the quantization values for each of the multiple blocks, the K parameter modeling correlation statistics of 5 the pixels in the image blocks and the C parameter modeling bits required to code overhead data.

35. The method of claim 34 including deriving the optimum quantization values in either a fixed mode where the 10 K and C parameters are known in advance or an adaptive mode where the K and C parameters are derived according to the K and C parameters of previously encoded blocks.

36. The method of claim 35 wherein the adaptive 15 mode includes the following steps:

- (a) deriving values for the K and C parameters that exactly predict the number of bits B used for encoding previous blocks;
- (b) deriving averages for the derived K and C parameters for the previously encoded video blocks; and
- (c) predicting the K and C parameters for a next video block by weighting the average K and C parameters according to the initial estimates 25 for the K and C parameters.

37. A method for encoding video comprising the 30 steps of:

- (a) detecting the location of a facial region of a frame of said video;
- (b) calculating a sensitivity value for each of a plurality of locations within said video based upon said location of said facial region; and

10  
5  
*Do not copy*

(c) encoding said frame in manner that provides a substantially uniform apparent quality of said plurality of locations to said viewer when said viewer is observing said facial region of said video.

10  
15  
20  
25  
30

38. The method of claim 37 wherein said sensitivity values are calculated based upon said location of said facial region, a size of said facial region, and a non-linear model of the sensitivity of a human visual system.

39. The method of claim 37 wherein said detecting said location of said facial region of said frame further comprises the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate regions within said difference image; and

(e) fitting said plurality of candidate regions to said difference image to select said facial region.

40. The method of claim 39 wherein said determining said plurality of candidate regions is based on a Hough transform of said difference image in a spacial domain to a parameter space.

41. The method of claim 39 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

5 42. The method of claim 41 wherein said threshold value is a predetermined value and said removing values is setting said values less than said threshold value to a selected value.

10 43. The method of claim 40 wherein said Hough transform is

$$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \quad \forall x_c, y_c, r \in (x - x_c)^2 + (y - y_c)^2 = r^2.$$

15 44. The method of claim 39 where said fitting of each of said candidate regions is based on a combination of at least two of the following three factors including, a fit factor representative of a fit of said candidate regions to said difference image, a location factor representative of the location of said candidate regions within said video, 20 and a size factor representative of the size of said candidate regions.

25 45. The method of claim 39 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

30 46. The method of claim 39 wherein said step of determining said plurality of said candidate regions and fitting said plurality of said candidate regions further comprises the steps of:

35 (a) determining a set of candidate circles based on a Hough transform of said difference image;

5 (b) scoring said set of said candidate circles based on a combination of at least three factors including, a fit factor representative of the fit of said candidate circles to said difference image, a location factor representative of the location of said candidate circles within said video, and a size factor representative of the size of said candidate circles;

10 (c) selecting at least one of said candidate circles based on said scoring;

(d) generating at least one candidate region having an elliptical shape for each of said at least one of said candidate circles; and

15 (e) scoring each of said candidate regions based on a combination of at least three factors including, a fit factor representative of the fit of a respective said candidate region to said difference image, a location factor representative of the location of said respective said candidate region within said video, and a size factor representative of the size of said respective said candidate region.

20

25                   47. The method of claim 46 wherein said  
generating at least one candidate region has a center of  
said elliptical shape located within a bounded region of  
potential locations having a greater vertical dimension than  
30 a horizontal dimension centered about the center of said  
respective said candidate circle.

48. The method of claim 38 wherein said wherein  
the step of said calculating said sensitivity values is  
35 further based upon calculating an eccentricity versus image

location in relation to a viewer of said video for said plurality of locations within said video.

49. The method of claim 48 wherein said eccentricity is derived according to the following,

10

15

20

where  $\theta_E$  is said eccentricity,  $y$  is a vertical pixel position within said video,  $x$  is a horizontal position within said video,  $x_c$  represents a horizontal component of a center position of an elliptical said facial region,  $y_c$  represents a vertical component of said center position of said elliptical said facial region,  $x_r$  represents a first elliptical radii of said elliptical said facial feature in a horizontal direction;  $y_r$  represents a second elliptical radii of said elliptical said facial feature in a vertical direction, and  $V$  represents a viewing distance.

25

50. The method of claim 38 wherein said sensitivity values are based upon the distance from the outer edge of said facial region to said plurality of locations within said video.

30

51. The method of claim 38 wherein said sensitivity versus eccentricity non-linear model is derived according to the following,

where  $S$  is representative of said sensitivity,  $k_{ECC}$  is a constant, and  $\theta_E$  is representative of a non-linear contrast sensitivity function.

5

52. The method of claim 37 further comprising the step of encoding said frame in manner that provides a substantially uniform apparent quality of said plurality of locations to said viewer when said viewer is observing said facial region of said video.

10

53. The method of claim 37 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

15

54. The method of claim 53 wherein said encoding is derived in accordance with the following:

$$Q/S_1, Q/S_2, Q/S_3, \dots, Q/S_N$$

20 where  $Q$  is representative of said base quantization factor, and  $S_1$  through  $S_N$  are representative of said sensitivity information for said plurality of locations.

25

55. The method of claim 54 wherein one of said  $S_k$ , where  $k$  is a value from 1 to  $N$ , is derived based upon a statistical calculation of a plurality of said sensitivity information for one of said locations of said image.

30

56. The method of claim 55 wherein  $S_k$  is an average of said plurality of said sensitivity information.

57. The method of claim 55 wherein  $S_k$  is a maximum of said plurality of said sensitivity information.

58. The method of claim 52 wherein said encoding said frame results in the total number of bits produced for said frame being substantially equal to a preselected number.

5

59. The method of claim 58 wherein said frame is encoded only once.

10 60. The method of claim 58 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

15

61. The method of claim 60 wherein said base quantization factor is derived in accordance with the following:

20

25

*Spn*  
*Att*

where  $A$  is representative of the number of pixels in one of said plurality of locations,  $K$  and  $C$  are constants associated with said plurality of locations,  $N$  is representative of the number of said plurality of locations,  $B$  is representative of said total number of bits, the  $\sigma_i^2$  values are a measure how much texture is associated with said plurality of locations, and the  $S_i^2$  values are representative of the respective said sensitivity information squared.